

Cambridge Language Sciences Symposium Poster Presentations

Tuesday 19 November 2019 – Magdalene College, University of Cambridge

- 1. A deep learning approach to automatic characterisation of rhythm in non-native English speech**
Konstantinos Kyriakopoulos, Kate Knill, and Mark Gales
- 2. A holistic measure of sociolinguistic experience: contextual and individual linguistic diversity in South Africa and the United Kingdom**
Mandy Wigdorowitz, Ana Pérez, and Ianthi Tsimpli
- 3. Automatic language identification in code-switched Hindi-English social media texts**
Sana Kidwai, Christopher Bryant, Li Nguyen, and Theresa Biberauer
**This research was supported by Cambridge Language Sciences Incubator Fund.*
- 4. An assessment of the knowledge, attitudes, and practices of slum-dwelling mothers pre- and post- ‘SODOTO’ model of intervention in the Kolkata Mobile Teaching Kitchen (MTK) project**
Sento Kai Kargbo, Luke Buckner, Minha Rajput-Ray, Maria Korre, and Sumantra Ray
**This research was supported by Cambridge Language Sciences Incubator Fund.*
- 5. Can you hear what’s coming? An ERP study of phonological prediction**
Victoria Poulton and Mante S. Nieuwland
- 6. Characterising musical interaction: an exploratory video analysis**
Rebecca Whiteman
- 7. Designing and developing a virtual English enrichment course for improving Chinese learners’ communicative competence in English**
Min Du
- 8. Greek accent in pronunciation of Russian coronal fricatives**
Daria Dashkevich
- 9. How are fact verification systems brittle to linguistic change?**
James Thorne and Andreas Vlachos
- 10. Learning (a language) at your brain’s pace**
Julia Heine, Elizabeth Michael, Zoe Kourtzi, Vicky Leong, Henriette Hendriks & John Williams
**This research was supported by Cambridge Language Sciences Incubator Fund.*
- 11. Neural computations of prediction error can explain MEG response during recognition of spoken words and pseudowords**
Yingcan Wang, Ediz Sohoglu, Rebecca Gilbert, Richard Henson, and Matthew Davis
- 12. Neural tracking of attended speech in monolingual and bilingual children**
Jacqueline Phelps and Mirjana Bozic

- 13. No, you're not alone: A better way to find people with similar experiences on Reddit**
Zhilin Wang, Elena Rastorgeuva, Weizhe Lin, and Xiaodong Wu
- 14. Postverbal subjects in Italian monolingual children**
Alexander Cairncross and Lena Dal Pozzo
- 15. Prosodic correlates of gapping in Mandarin Chinese**
Chenyang Zhang
- 16. Rapid evolution of the primate larynx**
Jacob Dunn
- 17. Referentiality and article errors in L2 English: evidence from Brazilian and Russian learners**
Kateryna Derkach and Theodora Alexopoulou
- 18. Script birth and script death: An interdisciplinary investigation**
Philippa Steele, Philip J. Boyes, Robert Crellin, Natalia Elvira Astoreca
- 19. The costs of faking it: A pilot study into the evolution of human accents in language**
Jonathan R Goodman, Francis Nolan, Chloe Allenby, and Robert Foley
** This research was supported by Cambridge Language Sciences Incubator Fund.*
- 20. The influence of linguistic and non-linguistic factors on second-language acquisition of tone**
Tim Laméris
- 21. Time for change: Evaluating models of semantic change without evaluation tasks**
Haim Dubossarsky, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg
- 22. Towards an automated event-semantic representation of L2 learner inflectional errors**
Carlos Balhana

1. A deep learning approach to automatic characterization of rhythm in non-native English speech

Konstantinos Kyriakopoulos, Kate Knill, and Mark Gales
Department of Engineering, University of Cambridge

A speaker's rhythm contributes to the intelligibility of their speech and can be characteristic of their language and accent. For non-native learners of a language, the extent to which they match its natural rhythm is an important predictor of their proficiency. As a learner improves, their rhythm is expected to become less similar to their L1 and more to the L2 they are learning.

Metrics based on the variability of the durations of vocalic and consonantal intervals have been shown to be effective at detecting language and accent. In this paper, pairwise variability, specifically PVI (Grabe and Low) and CCI (Bertinetto et al.), and variance metrics (Ramus et al.) are first used to predict proficiency and L1 of non-native speakers taking an English spoken exam.

Previous work (Kyriakopoulos et al.) showed that hierarchical attention-based recurrent neural networks could be used to generalise pronunciation features, producing a tunable end-to-end system for predicting grade and L1. In the same vein, a deep learning alternative to generalise the pairwise variability features is presented, in the form of a tunable duration embedding, based on attention over an RNN over durations. The recurrent layers allow relationships across time beyond just adjacent pairs to be captured, while self-attention allows sensitivity to the relative salience of different phones within each interval and different intervals within the utterance.

The systems are trained on a gender and proficiency level balanced mixed L1 dataset (TRN) consisting of 3376 speakers (first languages Polish, Vietnamese, Arabic, Dutch, French and Thai), scored on their overall proficiency by human graders and evaluated on a held-out evaluation set (EVL), consisting of 224 speakers of a similar mix of L1s, gender and proficiency, with scores provided by expert human graders.

2. A holistic measure of sociolinguistic experience: contextual and individual linguistic diversity in South Africa and the United Kingdom

Mandy Wigdorowitz, Ana Pérez, and Ianthi Tsimpli
Department of Theoretical and Applied Linguistics, University of Cambridge

Individual reports of language history, use, and proficiency are generally considered sufficient for language profiling; however, this neglects the influence of linguistic experience on linguistic knowledge. Linguistic experience refers to the significant inputs from the linguistic environment in which an individual has spent most of their life (their sociolinguistic context). Measuring linguistic experience can provide a holistic overview of where the speaker is linguistically immersed, including how certain languages may be privileged over others, and the speaker's fluidity and inter-mingling of language use. Moreover, such a measure would contribute to future research in language profiling. As yet however, the way in which sociolinguistic context facilitates linguistic gains has not been fully conceptualised or quantitatively investigated within the language sciences.

To address this, we develop and validate an evidence-based measure of linguistic experience – the Contextual and Individual Linguistic Diversity Questionnaire (CIL-D-Q) – and then use it to compare linguistic experience between participants from South Africa and the United Kingdom (UK). The following research questions guided the study: How many factors make up the structure of the CLID-Q as confirmed by an exploratory factor analysis? Are these factor-analysed scales reliable? Do speakers from South Africa and the UK differ in their contextual linguistic experience overall and across monolingual, bilingual, and multilingual groups?

Exploratory factor analysis was conducted with data from 353 participants (62.9% South African, 37.1% UK, $M = 29.3$, $SD = 10.09$). A three-factor solution was found to best describe the structure of the CILD-Q: Multilingualism in Context (contextual use and societal practice of multiple languages within a community), Multilingualism in Practice (direct and indirect linguistic exchanges and conversational interaction), and Linguistic Diversity Promotion (societal and governmental endorsement of linguistic variety). Item scores corresponding to these factors showed sufficient reliability with alpha coefficients > 0.80 . South African participants scored significantly higher than participants from the UK on the CILD-Q and across all language groups ($p < 0.05$).

The CILD-Q has been found to be a valid and reliable measure of contextual and individual linguistic experience. South African participants reported greater sociolinguistic experience than participants from the UK as a result of the language practices, communicative engagement, and encouragement from their society. Evidence supports the claim that the CILD-Q provides a novel and holistic way to measure linguistic experience and can be used when measuring linguistic variation across populations from differing sociolinguistic contexts, however, further studies are required to replicate these findings.

3. Automatic language identification in code-switched Hindi-English social media texts

Sana Kidwai, Christopher Bryant, Li Nguyen, and Theresa Biberauer

Department of Theoretical and Applied Linguistics, University of Cambridge

Natural Language Processing (NLP) tools often struggle to process code-switched data and so linguists are commonly forced to annotate such data manually. Automatic tools would nevertheless help speed up the annotation process and also improve consistency. Last year, Nguyen and Bryant developed a tool to semi-automatically annotate transcribed bilingual code-switched Vietnamese-English data with language and POS tags (Nguyen & Bryant, in submission). In this work, we attempt to extend their approach to another language pair, Hindi-English. Specifically, we applied their tool to data from the International Conference on Natural Language Processing (ICON) 2016 shared task, which consists of social media posts (Facebook, Twitter and Whatsapp) that have been annotated with language and POS tags (Molina et. al. 2016). We used the ICON-2016 annotations as the gold-standard labels in the language identification task.

Since Nguyen and Bryant originally developed their annotation tool for transcribed speech data, we first had to adapt it for social media data. In particular, we had to handle additional challenges such as misspellings, abbreviations, emoticons and URLs. Our modified system then made use of two monolingual word-lists and a manually defined

word map to assign three possible language tags: English (EN), Hindi (HI) and Universal (UNIV) (e.g. punctuation). Since there are also no standardised spellings for Hindi in the Roman Alphabet (Hindi social media users rarely write using the Hindi Devanagari script), we used a Hindi word-list that included several variants for each original Devnagari form. Whenever a word appeared in either both or neither word-list, we used a manually-defined word list to handle the most common cases. Ultimately, our tool achieved an accuracy of 90.68% on the ICON-2016 data. We then evaluated the first 500 tokens of each social media subset manually, and found 52.90% of all errors were caused by problems with the gold-standard (and correct identification by our system). Conversely, amongst tokens that had not been flagged as errors, only 0.95% were incorrectly identified by both the gold-standard and our system (false positives). It is thus likely that the overall accuracy of our system is higher than 90.68%. This shows significant potential for the extendibility of the Nguyen-Bryant tool in processing code-switched corpora, on different language combinations, and in different genres.

References:

- Nguyen, Li & Bryant, Christopher (In submission). CanVEC – the Canberra Vietnamese-English Code-switching Natural Speech Corpus: Recording, Transcription and Annotation.
- Molina, Giovanni, AlGhamdi, Fahad, Ghoneim, Mahmoud, Hawwari, Abdelati, Rey-Villamizar, Nicolas, Diab, Mona, Solorio, Tamar (2016). Overview for the Second Shared Task on Language Identification in Code-Switched Data. In Proceedings of the Second Workshop on Computational Approaches to Code Switching. Austin, Texas. Association for Computational Linguistics, 40 – 49.

4. **An Assessment of the Knowledge, Attitudes, and Practices of Slum-dwelling Mothers pre- and post- ‘SODOTO’ model of intervention in the Kolkata Mobile Teaching Kitchen (MTK) Project**

Sento Kai Kargbo, Luke Buckner, Minha Rajput-Ray, Maria Korre and Sumantra Ray

The Need for Nutrition Education/Innovation Programme (NNEdPro) Global Centre for Nutrition and Health in Cambridge, and Remedy Clinic Study Group in Kolkata and the Inner Wheel Club of Greater Calcutta launched the Mobile Teaching Kitchen (MTK) project as a nutritional education tool in two slums of Kolkata India in February 2018 to improve awareness of diet diversity and disease prevention by using locally sourced foods and cooking skills. Local volunteers trained in healthy cooking, transferred core principles through cooking demonstrations of sustainable, nutritional, and affordable meals. They followed a See One, Do One, Teach One’ (SODOTO) model using a storytelling approach in the primary language of local communication (with the language of key concepts constructed de novo rather than translated from materials written for Indian educators in English language) to transfer knowledge to their peers. Knowledge, attitudes and practices (KAP) of the participants were assessed pre- and post-intervention. Lastly, participants were assessed and required to fill out self-assessment questionnaires.

This study aims to investigate the relationship between demographic characteristics, namely, the education of the participants from the two slums (RG Kar Canal and Chetla), their occupation and income status, with their KAP and Teach One self-assessment.

Used primary data collected by NNEdPro in 2018 during the pilot phase of the mobile of

the mobile teaching clinics. Study includes 12 participants and final analyses include survey data. The ‘Teach One’ self-assessment and KAP survey responses were scored using a basic rating scale. Using Excel-based tools, measures of central tendency and variability were used to assess the impact of the SODOTO model of intervention in both slums as well as identify patterns between income, education, and occupation of the participants and their overall ‘Teach One’ assessment.

Participants had diverse demographic backgrounds. In the case of level of education, participants with some primary education (IV or higher) performed better during the ‘Teach One’ step with a mean score of 2.45 ($s=0.40$) than their counterparts who received no primary education (mean score =2.05, $s=0.19$). Additionally, women who were housewives on average scored higher at 2.36 ($s=0.17$) than their peers who were employed (mean=2.29, $s=0.50$). In the case of income level, women with low income (less than 2000/- a year) scored higher at an average of 2.4 ($s=0.20$) than their counterparts who earned between 2000/- and 4999/- (mean score =2.27, $s=0.39$) and above 5000/- (mean score = 2.33, $s =0.61$) respectively. Analysis of the KAP survey responses revealed overall improvements in knowledge, attitudes, and practices of the participants. Participants reported a mean score of 1.83 ($s=0.58$) post intervention which is higher than the mean score of 1.43 ($s=0.53$) reported pre-intervention suggesting that SODOTO model of intervention was effective in transmitting knowledge to participants. These results ought to be interpreted cautiously given the small sample size which may not be representative of the wider population, However this pilot work provides good direction to the types of data that may be collected and analysed in further work as this project is scaled over 2020/21.

Acknowledgements: Jodie Weber, Ianthi Tsimpli, Ananya Ria Roy, Lydia Akaje-Macauley, Marjorie Lima do Vale, Jorgen Johnsen, Matheus Abrantes and the Kolkata MTK Team

5. Can you hear what’s coming? An ERP study of phonological prediction

Victoria Poulton and Mante S. Nieuwland

MRC Cognition & Brain Sciences Unit, University of Cambridge

Prediction is proposed as a mechanism underlying the rapid processing of spoken language. However, the extent of these predictions is unclear. The strongest stances argue that predictions may include phonological information[1,2]. Previous studies describe two components of the event-related potential (ERP) – the N200[3] and the Phonological Mismatch Negativity (PMN)[4] – which are sensitive to unpredictable phonology. We investigated whether an “early” negative deflection in the ERP was (a) specifically sensitive to phonological violations and (b) distinct from the N400, which is associated with semantic processing [5,6].

Critical sentences such as “De bloemist bezorgde bij Lisa een mooi boeket/toetertje/tonnetje.” (English: the florist delivered to Lisa a beautiful bouquet/horn/barrel) were completed with a predictable word (boeket), a word which differed in the initial consonant but overlapped in phonology with the first predictable syllable (toetertje), or a word with no phonological overlap (tonnetje). We included unpredictable conditions with different degrees of overlap in order to test whether ERP

amplitude is modulated by overlap or instead reflects an “all-or-none” mismatch detection process [7].

EEG data from 24 participants has been analysed. We examined the average voltage at sentence-final words in early (150-250ms post-stimulus) and late time windows (300-500ms post-stimulus, typical of the N400). Voltage was averaged across electrodes in anterior and posterior regions of interest (ROI).

The waveforms of the unpredictable conditions diverged from the predictable condition as early as 150ms post-stimulus. Linear mixed-effects regression revealed a more negative amplitude for unpredictable compared to predictable words, while pairwise comparisons of the least-squares means showed that the unpredictable conditions did not differ from each other. To examine scalp distribution, we conducted a 2x2 ANOVA, with main effects of ROI and time window, on the z-transformed voltage difference between the no-overlap and predictable conditions. The absence of significant main effects or an interaction, further qualified by a Bayes factor analysis, suggests that there were no significant distributional differences between the two time windows. This conflicts with reports of a distinct component and instead supports the interpretation of these effects as an early onset to the N400[8]. As we have no evidence for a phonology-specific N200/PMN, the findings cannot support the strongest stances on prediction. It is possible that ERPs are not suitable to detect true phonological mismatch effects during sentence comprehension. However, the nature of the observed effect is consistent with evidence for the early influence of predictions generated from semantic contextual constraint during word recognition[9].

References:

1. Altmann, G. T. M. & Mirković, J. Incrementality and Prediction in Human Sentence Processing. *Cogn. Sci.* 33, 583–609 (2009).
2. Pickering, M. J. & Gambi, C. Predicting while comprehending language: A theory and review. *Psychol. Bull.* 144, 1002–1044 (2018).
3. Van den Brink, D., Brown, C. M. & Hagoort, P. Electrophysiological Evidence for Early Contextual Influences during Spoken-Word Recognition: N200 Versus N400 Effects. *J. Cogn. Neurosci.* 13, 967–985 (2001).
4. Connolly, J. F. & Phillips, N. A. Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *J. Cogn. Neurosci.* 6, 256–266 (1994).
5. Kutas, M. & Hillyard, S. A. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163 (1984).
6. Kutas, M. & Federmeier, K. D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* 62, 621–647 (2011).
7. Newman, R. L., Connolly, J. F., Service, E. & Mcivor, K. Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology* 40, 640–647 (2003).
8. Van Petten, C., Coulson, S., Rubin, S., Plante, E. & Parks, M. Time course of word identification and semantic integration in spoken language. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 394–417 (1999).
9. Klimovich-Gray, A. et al. Balancing Prediction and Sensory Input in Speech Comprehension: The Spatiotemporal Dynamics of Word Recognition in Context. *J. Neurosci.* 39, 519–527 (2019).

6. Characterising musical interaction: an exploratory video analysis

Rebecca Whiteman

Faculty of Music, University of Cambridge

When we hold a conversation, we work together in a process of grounding, sharing common understanding about something (Clark, 1996), within a phatic channel of engagement (Coupland et al., 1992). In music, we foreground the phatic dimension over transactional goals, creating a context of floating intentionality (Cross, 2005, 2013). So far, social interaction research has focussed on speech: its turn-taking structure (Sacks et al., 1978), patterns of gaze (Kendon, 1967) and synchrony (Condon & Ogston, 1967) which together support the grounding process. This exploratory video case study investigated characteristics of musical interaction. I compared simple interaction tasks, with and without an external goal (more like speech or music, respectively), analysing patterns of gaze behaviour and synchrony.

Six participants (aged 19-22) took part, tested in pairs (participants had no prior interaction). Each pair was instructed to ‘make music’ together for five minutes, using a simple instrument (no speech). One pair had no extra goal; one was to depict a flock of birds; and the other a herd of elephants (no goal, an easier and a harder goal, respectively). Participants consented to be filmed and their individual contributions recorded using personal microphones. I identified periods of joint synchrony from audio data (using Praat), analysed gaze behaviour from video footage, and investigated the relationship between gaze and patterns of turn-taking (using ELAN).

I found longer periods of joint synchrony in the no-goal (78% of interaction time) than the goal conditions (3.9% and 1.9% in bird and elephant conditions). There were also distinct gaze and turn-taking patterns. In the no-goal condition, participants played simultaneously throughout, co-ordinating at phrase transitions; participants looked towards the other (and mutual gaze occurred) just before or during these moments. Participants with the bird goal played in turns; each looked away before and then checked back to the other (achieving mutual gaze) after the onset of their turn. Participants given the harder, elephant goal displayed a mixture of these two patterns.

In summary, when an external goal is salient (or easy), participants displayed the turn-taking and gaze aversion-gaze checking patterns characteristic of speech (compare e.g. Bavelas et al., 2002). However, where the goal was harder or where there was no extra goal (more “musical” forms of interaction), participants look to the other as they co-ordinate phrase transitions, and spend more time in synchrony. These initial findings show clear behavioural differences between interaction which is more like speech or music.

References:

- Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52(3), 566—580. <https://doi.org/10.1111/j.1460-2466.2002.tb02562.x>
- Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press (ACLS Humanities E-book)

- Condon, W. S., & Ogston, W. D. (1967). A segmentation of behavior. *Journal of Psychiatric Research*, 5, 221–235. [https://doi.org/10.1016/0022-3956\(67\)90004-0](https://doi.org/10.1016/0022-3956(67)90004-0)
- Coupland, J., Coupland, N., & Robinson, J. D. (1992). “How are you?”: Negotiating phatic communion. *Language in Society*, 23, 207–230.
- Cross, I. (2005). Music and meaning, ambiguity and evolution. In D. Miell, R. MacDonald, & D. Hargreaves (Eds.), *Musical Communication* (pp. 27–44). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198529361.003.0002>
- Cross, I. (2013). “Does not compute”? Music as real-time communicative interaction. *AI and Society*, 28, 415–430. <https://doi.org/10.1007/s00146-013-0511-x>
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26, 22–63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>

7. Designing and developing a virtual English enrichment course for improving chinese learners’ communicative competence in English

Min Du

Faculty of Education, University of Cambridge

Despite the significant emphasis on English learning in China, the current teaching strategies and pedagogies often fail to prepare English learners to cope with international life and the complex communication scenarios in English. To improve this situation, this study aimed to design and implement an online virtual English Enrichment Course (EEC) to develop Chinese English learners' communicative competence and explore learners' perceptions of the course features that contributed to developing their communicative competence.

The development of the course was a pioneering empirical study based on the complex educational design theory and supported by the communicative language teaching theory, learning theories (complex dynamic systems theory and Vygotsky’s sociocultural theory) and conversation-based communicative computer-assisted language learning. The EEC course adopted an online language course design model aiming to respond to the unpredictability and instability of the constantly changing real world. The topics included up-to-date news items and cultural rituals in England. Invited English speakers offered topics in which they were specialised. Student presentation followed by Question and Answer time was also introduced as novel in the Chinese context.

As design-based research, this study comprised four cycles of development, implementation, analysis and refinement of the course over four years. Eight to ten 16- to 21-year-old students from China attended the EEC course in each research cycle; Research Cycles 3 and 4 additionally involved comparison groups who undertook their usual English language teaching. The effectiveness of the course was measured by comparing students' progress on pre-tests and post-tests. The assessment criteria for linguistic competence and part of interactional competence employed the assessment framework produced by Cambridge English. However, the assessment criteria for sociocultural competence was developed through this study. Students' perceptions of the EEC course were examined through questionnaires and interviews.

The findings showed that the EEC course significantly helped students to improve the sociocultural and interactional elements of communicative competence. Key features identified by the students as helping them to improve their communicative competence included the interactive nature of the course, the specialised knowledge of the invited English speakers concerning the topics, and the intercultural communicative competence orientation. A key contribution of the study is that the EEC course demonstrated a practical approach to applying and integrating CLT (Communicative language teaching) to communicative CALL (Computer-Assisted Language Learning), which allows more authentic English communication to take place across the globe. Moreover, this study not only offered an empirical investigation pertaining to the complex educational design model but also developed its theory.

8. Greek accent in pronunciation of Russian coronal fricative

Daria Dashkevich

Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge

This work presents some challenges and mispronunciations typical for native speakers of Greek language, who study the Russian language, in regards to Russian coronal fricatives. The Russian phonetic system is wider than the Greek in regards to fricatives. That is why we could predict some challenges in pronunciation of Russian fricatives by Greek speakers. In order to study this topic, we ran a linguistic experiment. First of all, we made predictions about the main problems that may appear in the pronunciation of the Russian sounds by the speakers of Greek language. Next, we made a focused reading materials in Russian, based on our predictions, and we gave these materials to the students, who's native language was Greek. Finally, we analysed the recordings, marking out typical problems for Greek students in learning Russian phonetics in regards to coronal fricatives.

Having compared the phonetical systems of Russian and Greek languages in the domain of coronal fricatives we could predict several problematic areas for Greek students learning Russian:

- Pronunciation of Russian soft and hard consonants;
- Differentiation between Russian sounds [s] – [z] and [ʃ] – [ʒ], [ɛ:];
- Pronunciation of voiced affricate [dʒ] instead of voiceless [ts] in several positions;
- Incorrect pronunciation of Russian voiceless alveolo-palatal sibilant affricate [tʃ];
- Incorrect pronunciation of Russian voiceless alveolo-palatal sibilant fricative [ç].

In order to verify our expectations, we gave special focused reading materials to several Greek students, who had different level of Russian.

Firstly, Greek students, indeed, found it difficult to differentiate between Russian hard and soft consonants. For example, they pronounced *chu[t] instead of chu[tʲ] (чусть), *[sʲ]yn instead of [sʲ]yn (сын) and other.

Secondly, lots of mistakes related to differentiation of [s]/[sʲ] – [z]/[zʲ] and [ʃ] – [ʒ], [ɛ:] were found in the speech of respondents, such as *[z]ournalchik instead of [zʲ]ournalchik (журнальчик), *sme[s]no instead of sme[ʃ]no (смешно), *be[ɛ:]apki instead of be[ʃ:]apki (без шапки), *kra[ɛ:]ivyh instead of kra[sʲ]ivyh (красивых).

Here are some other problematic areas for Greek speakers found in the study:

- pronunciation of sound combinations [stɐ] or [stʲɐ] instead of voiceless retroflex fricative [ɕ];
- pronunciation of affricates instead of [ʂ] and [ɕ:];
- differentiation between voicing of voiceless sounds;
- pronunciation of soft [tʲ] instead of affricate [tɕ] and others.

Thus, in this research we marked out the main problems in acquisition of Russian coronal fricatives by speakers of Greek language. The results of this paper may be implemented in Russian classes to guide pronunciation of students, whose native language is Greek.

Bibliography:

1. Barkhudarova E.L. Methodological problems in the analysis of the foreign accent of Russian speech // Gazette of Moscow University. Series 9. Philology. - 2012. - № 6. P. 57 -70.
2. Reformatskiy A.A. Phonology in service of teaching pronunciation of non-native language // Reformatskiy A.A. From the history of Russian phonology: Essay; Reader. - M., 1970. - P. 506 - 515.
3. Scherba L.V. Teaching of foreign languages in high school: General methods. - M., 1974. - 111 p.

9. How are fact verification systems brittle to linguistic change?

James Thorne and Andreas Vlachos

Department of Computer Science and Technology, University of Cambridge

Fact verification is the task of predicting whether claims can be supported or refuted by evidence. Advances in this task have been achieved through improved modelling and the availability of resources to train and validate systems. As this has potentially sensitive applications, like propaganda or biased news detection, it is critical to understand how systems and models behave when exposed to real-world data and how deficiencies in their training data may contribute to this. It has been observed in related NLP tasks that as models become more complex, it is difficult to fully understand and characterize their behaviour. And from an NLP perspective, there has been an ongoing discussion as to what extent these models understand language or they are exploiting unintentional biases and cues that are present in the datasets they are trained on.

To investigate how these biases manifest, we employ adversarial evaluation: where instance designed to induce classification errors are use to expose “blind spots” of systems under test. We consider six fact verification systems from the recent Fact Extraction and VERification (FEVER) challenge which share a common architecture. These systems comprise models that retrieve evidence, aggregate it, and classify whether a claim is supported or refuted is supported refuted given this evidence.

We demonstrate deficiencies in fact verification systems and show that errors can be induced through meaning-preserving linguistic change using both automated and manual methods that exploit model behaviour, lexical resources and dataset statistics.

We compare which types of variation have the largest reductions in accuracy on the models under test, highlighting the most brittle components. These findings will inform

both design decisions for fact verification systems as well as future directions for datasets to make systems more robust to linguistic change.

10. Learning (a language) at your brain's pace

Julia Heine, Elizabeth Michael, Zoe Kourtzi, Vicky Leong, Henriette Hendriks & John Williams

Department of Theoretical & Applied Linguistics / Department of Psychology, University of Cambridge

All learners are different, showing different ways of tackling material to be acquired. But to date we understand very little of individual learner differences, and how to best adjust teaching to their needs.

Here, we aim to shed further light on individual (language) learning, by making use of neuroscientific research on the rhythm at which the brain processes information: Recent research has shown that neuronal oscillations (rhythmic and/or repetitive electrical activity generated spontaneously or in response to stimuli by neuronal populations in the central nervous system) predictively track rhythmic patterns in the environment, to optimally parse and extract temporal structure as, for example, in spoken language. Consequently, adults performed better at both basic perceptual and complex sentence-processing tasks when stimuli were delivered at a regular, predictable pace.

Research has further shown, however, that there are large individual differences in terms of the base frequency at which neuronal networks prefer to operate. Thus, although the average peak Alpha EEG frequency (the relevant measure of neuronal oscillations) across individuals is approximately 10 Hz, about 30% of individuals will have peak Alpha frequencies that are higher or lower, with the peak rate of the slowest individuals (~6 Hz) half that of the fastest. Delivery at one regular, predictable pace may therefore be beneficial, but will not necessarily be optimal for every learner. If, however, one could match the rate of stimulation in a learning situation with the individual's spontaneous oscillation rate, this might improve the efficacy of learning.

Here we propose to test this hypothesis by aligning the pace of teaching of an artificial language (statistical learning task) with the endogenous oscillatory rate of individual participants, to see if this personalization is more effective for learning than delivery at a group mean rate. Using language as the taught material is particularly relevant as spoken language is strongly rhythmically-patterned at different timescales, and it is known that this temporal structure drives entrainment of neuronal oscillations. Further, neural entrainment of neuronal oscillations by speech temporal structure is crucial for speech comprehension.

Although there is strong evidence supporting the importance of brain rhythms in neural information processing, whether these mechanisms can be exploited to optimize learning has not yet been tested, but cf. current work by Kourtzi and Leong. Testing the hypothesis in the current project will advance our understanding of the brain mechanisms that underlie individual learning ability and start to translate this knowledge into personalised interventions, something eagerly awaited by learners and (language) educators alike.

11. Neural computations of prediction error can explain MEG response during recognition of spoken words and pseudowords

Yingcan Wang, Ediz Sohoglu, Rebecca Gilbert, Richard Henson, and Matthew Davis
MRC Cognition & Brain Sciences Unit, University of Cambridge

How do human brains process spoken words? A phenomenon that helps us understanding spoken word recognition is the competitor priming effect (Monsell & Hirsh, 1998), i.e. hearing the word /haidzæk/ delays the recognition of the competitor word /haidzi:n/, which begins with the same initial segments. Models supporting competitive selection (e.g. TRACE, McClelland & Elman, 1986) propose that this is due to lateral inhibition caused by lexical competition, which makes competitively primed word /haidzi:n/ less activated than before. Whereas predictive selection accounts (e.g. Predictive Coding (PC), Davis & Sohoglu, in press) suggest that because of prior presentation of /haidzæk/, segment prediction between /haidz/ and /æk/ was enhanced, hence on hearing /haidzi:n/, greater prediction error would be evoked, which in turn delays its recognition. Here, we report an MEG study that distinguishes these accounts.

Twenty-four participants made lexical decisions on 160 sets of spoken words (e.g. captain) and pseudowords (e.g. captick) either with or without prior auditory presentation of another item sharing the same initial segments (e.g. hearing captive before captain or captick). While performing this task, participants' brain activity was recorded by MEG. Neural responses were time-locked to the deviation point (DP) where stimuli acoustically diverge from each other (e.g. immediately after /ɪ/ for /kæptɪ.../).

Behavioural results replicated the competitor priming effects, where word recognition was delayed by prior presentation of a similar word but not pseudoword. MEG results indicated that from around 400-700ms post DP, pseudowords elicited significantly stronger neural responses in the superior temporal gyrus (STG) than words, and in line with behavioural effects, neural signals evoked by words and pseudowords showed an interaction between lexicality and prime type. Specifically, competitively primed words evoked stronger neural responses in the STG than unprimed words, but neural signals evoked by primed and unprimed pseudowords did not differ.

Competitive selection accounts propose that the degree of lexical uncertainty correlates with the strength of neural responses in inferior temporal and frontal regions, and lexical uncertainty should be minimal after the DP when one or no word can be selected regardless of different prime types. In contrast, the PC account suggests that the size of PE correlates with neural responses in the STG, and larger PE, hence neural responses, should be elicited after competitor priming or by pseudowords due to the violation of prediction post-DP. These predictions based on the PC account are more in line with our MEG results.

12. Neural tracking of attended speech in monolingual and bilingual children

Jacqueline Phelps and Mirjana Bozic
Department of Psychology, University of Cambridge

Bilingualism has been shown to modulate the neural mechanisms of selective attention, with differences between monolinguals and bilinguals observed even when they display equivalent behavioural performance (e.g., Olguin et al, 2019). This suggests that the

crucial consequence of learning and using multiple languages might be that it triggers neuroplastic adaptation that allows bilinguals to achieve optimal performance under increased processing demands, but which is more subtle than what is captured by behavioural tests alone. To date, these studies have focused primarily on adults and it is therefore unclear to what extent such changes affect the mechanisms of selective attention in bilingual children. One possibility is that the demands of competition and inhibition between co-activated languages reconfigure the patterns of attentional processes right from the onset, such that the effects can be clearly discerned by the time children can respond to selective attention tasks. Alternatively, these modifications might have a protracted maturation dependent on the length of exposure to the demands of bilingualism, in which case they would be much more evident in adults than in children.

To address this question, we tracked the neural encoding of attended continuous speech in 24 monolingual and 24 bilingual children aged 7-12, in the context of different types of acoustic and linguistic interference. Participants always attended to a narrative in English while four different types of interference were presented to the unattended ear: a different narrative in English, a narrative in Latin (a language unknown to the listener), a well-matched non-linguistic acoustic interference (Musical Rain), and no interference. The neural activity was recorded by a 64-channel EEG system and cross-correlated with the speech envelopes for both attended and unattended streams. Results showed more robust neural encoding for the attended envelopes than ignored ones across all conditions for both bilinguals and monolinguals. We also saw that the type of interference significantly modulated the encoding of attended speech for monolinguals. Bilingual children on the other hand displayed equivalent encoding of the attended stream in every condition, directly replicating the pattern of results previously observed in adults. These findings indicate that, even when behavioural responses are equivalent, monolingual and bilingual children exhibit different patterns of neural entrainment to attended speech envelopes. This suggests that the reconfiguration of the neural mechanisms of selective attention in bilinguals is established by the time children can reliably respond to selective attention tasks.

13. No, you're not alone: A better way to find people with similar experiences on Reddit

Zhilin Wang, Elena Rastorgeuva, Weizhe Lin, and Xiaodong Wu

Faculty of Education, University of Cambridge

We present a probabilistic clustering algorithm that can help Reddit users to find posts that discuss experiences similar to their own. This task is important because doing so can help users understand how others respond to issues in their lives similar to theirs and therefore learn to respond more effectively to their own situations. Reddit is chosen as a source of written posts because Reddit users do not typically use real names, which provide a sense of anonymity. This means that users might be more willing to share their experience without facing judgment by people who they know in real life.

This model is built upon the BERT Next Sentence Prediction model. BERT Next Sentence Prediction model is chosen because it is shown to demonstrate good performance in a variety of sentence/paragraph-level semantic tasks by better understanding the complex non-linear relationships between two text rather than depending on a distance metric between the embedding of two text only.

Our innovation reduces the time complexity for clustering all posts in a corpus from $O(n^2)$ to $O(n)$ with respect to the number of posts. When evaluating using the proxy task of similarities between texts in the same cluster, we demonstrate that such probabilistic clustering can yield a performance better than baseline clustering methods based on Latent Dirichlet Allocation (Blei et al., 2003) and Word2Vec (Mikolov et al., 2013). Furthermore, there is a high degree of coherence between our probabilistic clustering and the exhaustive comparison $O(n^2)$ algorithm in which the similarity between every pair of posts is found. This makes the use of the BERT Next Sentence Prediction model more practical for unsupervised clustering tasks due to the high runtime overhead of each BERT computation.

14. Postverbal subjects in Italian monolingual children

Alexander Cairncross and Lena Dal Pozzo

Theoretical and Applied Linguistics Section, University of Cambridge

The use and interpretation of preverbal null and overt sentential subjects in null-subject languages has been widely investigated in language acquisition. It is well-known that while the syntactic option may be early acquired, convergence with the target on the discursive conditions for null and overt subjects is more demanding in various populations (e.g. Sorace et al. 2009, Belletti et al. 2007). Another less frequent but related feature of Italian style null-subject languages is the occurrence of postverbal subjects. In Italian these postverbal subjects may be due to argument structure (unaccusatives) or discourse structure (e.g. new information focus). The latter differ both in their syntactic position and crucially in their need for the integration of discourse information (Burzio 1986, Belletti 2001 a.o.). Focusing on Italian, monolingual children have acquired postverbal subjects in unaccusatives by the start of school (Lorusso et al. 2004, Lorusso 2006, Vernice and Guasti 2014). To date however, little research has explored Italian postverbal subjects in new information focus contexts in child acquisition (cf. Dal Pozzo 2012).

This work therefore addresses the production and comprehension of Italian subjects under three conditions: (a) broad focus (Che cosa è successo? What happened?), (b) new information focus on the subject (Chi è arrivato? Who has arrived?), (c) new information focus on the event (Cosa fa X? What did X do?), which was included for comparison with work on overt preverbal subjects (cf. Sorace and Serratrice 2009 a.o.). We expect that: (i) preverbal subjects are preferred when answering questions of type (a), postverbal subjects for questions of type (b) and null subjects for questions of type (c). Participants were monolingual children (N=34; aged 6;1-7;1). Data was collected through an elicitation task and a forced choice task. The elicitation task consisted of 2 sets of 25 pictures for which the children answered 30 questions of the type in (a), (b), and (c) above. The forced choice task consisted of 40 scenes in which Mickey Mouse and Donald Duck responded to prerecorded questions of the types above. Children were required to indicate “who responded better,” where responses differed only by subject position.

Results indicate children prefer postverbal subjects in narrow focus contexts (ET and FCT $ps < 0.00$). Results also replicated the previously reported unaccusativity effect in broad focus (FCT $p < 0.05$). The preference for post verbal subjects in narrow focus is independent of unaccusativity. Moreover, postverbal subjects are more frequent with transitive verbs than unaccusatives and unergatives (FCT $ps < 0.05$). In conclusion, our

evidence shows that monolingual children by their first year of school have acquired postverbal subjects in narrow focus, understand their discursive distribution, and differentiate these subjects from postverbal subjects in unaccusatives.

15. Prosodic correlates of gapping in Mandarin Chinese

Chenyang Zhang

Faculty of Asian and Middle Eastern Studies, University of Cambridge

An experiment is reported which tests the prosodic reflexes of verb-less constructions in English and Mandarin Chinese (henceforth, MC). The prosody is predicted to reflect the differences in the underlying syntactic structures between gapped and non-gapped counterparts in the two languages. This experiment constitutes part of the empirical study on the second language acquisition by English-speaking learners of verb-less constructions in MC of the kind illustrated in (1):

(1) Chinese gapping sentences:

Zhangsan chi-le san-ge-pingguo, Lisi chi-le si-ge-juzi.

Zhangsan eat-ASP 3-CL-apple Lisi eat-ASP 4-CL-orange

‘Zhangsan ate three apples, and Lisi four oranges.’ (ASP: aspect marker; CL: classifier)

The Chinese surface forms look like canonical gapping in English, as seen in the English translation in (1), but the constructions are not parallel. There are semantic and discourse constraints unique to Chinese gapping sentences which do not apply with canonical English gapping. For instance, when the object in the gapped conjunct can function as an individual-level predicate (referring to some generic, permanent property or characteristic, like nationality, animate proper name, etc), the gapping sentence is unacceptable in MC:

(2) *Women xihuan meiguoren, tamen xihuan deguoren.

We like Americans they like Germans

‘We like Americans, and they Germans.’

However, the acceptability in MC can be improved in the context of an appropriate precursor:

(3) Q: Nimen xihuan naguoren?

You like which-country-people

‘Which country’s people do you like?’

These constraints on gapping sentences might be attributable to the ambiguity which arises from the existence of another type of verb-less constructions in MC, noun predicate sentences:

(4) Zhangsan zhongguoren.

Zhangsan Chinese

‘Zhangsan is Chinese.’

In noun predicate sentences, the subject and the noun predicate directly compose a small clause without projecting to any VP, which is associated with an identity reading.

Therefore, Chinese sentences with an individual-level predicate nominal and no overt verb have two possible readings:

(5) Gapping reading: ‘He likes Germans, and I like Americans.’

Identity reading: ‘He like Germans and I am American.’

The identity reading is always salient in ‘unmarked’ contexts.

Although the surface forms of the gapping and noun predicate constructions in MC are the same, the difference in underlying structure means that there is a stronger syntactic boundary between the nominals in the gapped conjunct. It is hypothesised here that this may result in prosodic features marking the pre-gap nominal such as (a) pre-boundary lengthening and (b) (for MC) the blocking of tone 3 sandhi. In the experiment to be reported, participants (native MC and English) read out a list of gapping sentences and non-gapping sentences in different contexts. Preliminary results provide support for (a) but not (b). The full results will be presented and discussed.

16. Rapid evolution of the primate larynx

Jacob Dunn

Department of Archeology, University of Cambridge

Understanding the evolution of vocal communication requires detailed knowledge of the sound production mechanism, which in mammals centres on the larynx. The mammalian larynx is a key target for selection, particularly in species with highly developed auditory-vocal communication. Identifying the resulting adaptive modifications depends on understanding laryngeal morphology in phylogenetic context. Using state-of-the-art phylogenetic comparative methods and a novel database of digital measurements obtained from 3-dimensional CT scans, we compare and contrast laryngeal specimens from 26 primate and 29 carnivoran species, covering nearly the full body size range in both orders. Our results show that the primate larynx has evolved more rapidly than the carnivoran larynx, resulting in a pattern of larger overall size and increased deviation from expected allometry with body size. We also show that residual larynx size is a strong predictor of call fundamental frequency in our sample, underscoring the relevance of our results for the evolution of vocal communication. Using excised larynx experiments, we show that, across species, vocal fold length predicts minimum fundamental frequency much better than body size, clearly demonstrating the potential for decoupling between larynx size and body size. Overall, these results imply that selective pressures on the mechanistic determinants of vocalization may be especially strong in primates, consistent with the hypothesis that the complexity of primate social organization is a key driver underlying vocal complexity. Such inference may provide insight into mechanisms underlying vocal complexity and the evolution of speech.

17. Referentiality and article errors in L2 English: evidence from Brazilian and Russian learners

Kateryna Derkach and Theodora Alexopoulou

Department of Theoretical and Applied Linguistics Section, University of Cambridge

We present an ongoing learner corpus analysis of article use in English L2 by learners with typologically distinct languages (currently L1-Brazilian (i.e. Brazilian Portuguese))

and L1-Russian). Employing large learner corpora allows to explore a variety of contexts and discover patterns of article use. We analyse article use and errors to identify learner challenges and consider their possible causes. A multitude of features plays a role in learners' choice of articles. Learners from article-less languages may misanalyse the definite article as a specificity marker (Ionin, Ko, & Wexler, 2004) or may omit articles for salient referents (Trenkic & Pongpaiboj, 2013), while types of discourse-pragmatic contexts from Hawkins' taxonomy (1978) also interact with article use (Ionin & Montrul, 2010; Liu & Gleason, 2002; Robertson, 2000). Article choice is also possibly influenced by noun phrase (NP) features, such as number, count/mass distinction, and abstractness (Snape, 2008). We draw data from EFCAMDAT, a large written learner corpus, which spans across 16 proficiency levels and 128 distinct tasks. The exploratory analysis involves 160 scripts (80 L1 Brazilian – article language, 80 L1 Russian – article-less, A2-B1 CEFR). We manually coded 1546 NPs for obligatory contexts: indefinite (n=394; 25%), definite (n=584; 38%) or obligatory bare (n=568; 37%), the latter being largely overlooked in article error research. The data was coded for: (i) error types: omission (n=183; 67% of all errors), oversuppliance (n=50; 18%), substitution (n=42; 15%); (ii) discourse-pragmatic and semantic uses (e.g. existential, anaphoric, specific/non-specific) combining Hawkins' taxonomy and the notion of specificity (Ionin et al., 2004; Trenkic, 2008), as well as (iii) NP type: singular count / plural count / mass; abstract / concrete. The preliminary results show an overall accuracy rate of 87% for Brazilians and 78% for Russians, but there are significant differences across conditions. L1-Russian learners are much less accurate than L1-Brazilian learners on obligatory article contexts (63% vs. 81% on a/an, 72% vs. 89% on the), while being more accurate on obligatory bare contexts (94% vs. 88%). Crucially, they differ in their errors: Brazilians oversupply the (63% of errors in obligatory bare condition), while Russians oversupply both articles. All learners struggle with non-referential uses of the indefinite article (e.g. I will find a good job; these constitute over 50% of obligatory indefinite contexts): Russians' accuracy is at 54%, and Brazilians' at 79%. Our hypothesis is that learners link the use of the article with the referentiality of the nominal, rather than definiteness, while they have difficulty grasping the role of grammatical number in the choice between the indefinite article and the zero article and its interaction with the count/mass distinction. Specifically, learners assume that the zero article marks non-referential nominals resulting in much higher error rates in this condition and higher omission with abstract nouns.

References:

- Hawkins, J. A. (1978). *Definiteness and Indefiniteness*. London: Croom Helm.
- Ionin, T., Ko, H., & Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1), 3–69.
- Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60(4), 877–925.
- Liu, D., & Gleason, J. L. (2002). Acquisition of the article “the” by nonnative speakers of English: An Analysis of Four Nongeneric Uses. *Studies in Second Language Acquisition*, 24, 1–26.
- Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 16(2), 135–172.
- Snape, N. (2008). Resetting the Nominal Mapping Parameter in L2 English: Definite article use and the count–mass distinction. *Bilingualism: Language and Cognition*, 11(01), 63–79.
- Thomas, M., (1989), The acquisition of articles by first and second language learners, *Applied Psycholinguistics*, 10(03), 335-355.

- Trenkic, D., & Pongpaibroj, N. (2013). Referent salience affects second language article use. *Bilingualism: Language and Cognition*, 16(1), 152–166.
-

18. Script birth and script death: An interdisciplinary investigation

Philippa Steele, Philip J. Boyes, Robert Crellin, Natalia Elvira Astoreca
Faculty of Classics, University of Cambridge

Poster presentation, proposed by the members of the project ‘Contexts of and Relations between Early Writing Systems’ (CREWS).

Script changes are a recurring theme of the CREWS project, which unites researchers working on Aegean and Near Eastern writing systems from multiple disciplinary perspectives. For the ancient world, we can only access language through writing, and so the systems used by ancient societies to encode their languages have a significant impact on our ability to access and analyse their linguistic features. In this investigation we are particularly interested in the motivations for the creation of writing systems (often referred to as scripts) and their eventual loss in some societies, processes that affect our access to linguistic information but that are also linked with episodes of linguistic and social (and sociolinguistic) change. ‘Script death’ is arguably an unhelpful label for the disappearance of a writing system, a process typically marked by significant social change affecting the core spheres in which writing operated (examples: Linear B, Ugaritic cuneiform). Investigating the social role of writing is therefore crucial to understanding why a script was discontinued. Equally, ‘script birth’ usually does not involve ex novo creation, but rather borrowing writing from elsewhere, involving adaptive processes that are social and contextual as much as linguistic (examples: Greek alphabet, Phoenician, Punic). This wider comparative perspective identifies common factors of script change and emphasises the cultural background to writing traditions, spanning the eastern Mediterranean and Levant. Crucially, we highlight the importance of holistic, integrated approaches that do not separate but rather interweave epigraphic, linguistic, archaeological and anthropological research. We also demonstrate contemporary relevance of script change, which plays a pedagogical role in our outreach activities. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 677758).

19. The cost of faking it: a pilot study into the evolution of human accents in language

Jonathan R Goodman
Leverhulme Centre for Human Evolutionary Studies, University of Cambridge

The Red Queen model suggests that evolution is a persistent competitive race between lineages, each trying to stay ahead of the game (eg, arms races, predator-prey relations, hosts and parasites). If language is conceived as a competitive interaction between signaller and listener, then a comparison to an evolutionary arms race may provide insights into how that competitive interaction may be stopped, slowed, or redirected. We might expect, for example, that if there are any linguistic mechanisms for signalling kinship affiliation, that some fakers will become adept at feigning those signals effectively, which will in turn lead to mechanisms for detecting fakes. Over time, this might help to explain the evolution of the complexity of accents of human languages.

This leads us to make several rudimentary hypotheses about the evolution of accents: first, that people will be adept at determining when speakers are mimics, particularly when the mimicked accent is familiar; second, that a speaker's ability to mimic correlates positively with the speaker's familiarity with accents of English, and with the number of languages the speaker has knowledge of.

For this pilot study, we are recruiting 50 participants to attempt to mimic accents of English with which they may be unfamiliar, to attempt to determine whether several recordings are those of mimics or genuine speakers, and finally to complete a short questionnaire about their own background. Here we present our initial results of this ongoing study.

20. The influence of linguistic and non-linguistic factors on second-language acquisition of tone

Tim Laméris

Department of Theoretical and Applied Linguistics Section, University of Cambridge

Unlike English, most of the world's languages use tone to change a word's core meaning. Learning tone in a second language (L2) is known to be difficult, but previous research has shown that both linguistic factors (such as native language experience with tone) and non-linguistic factors (such as musical experience) can make learning tones easier for some learners [1]. However, most previous studies have only separately examined either linguistic or non-linguistic influences on L2 tone acquisition, with a few exceptions, such as [2], [3]. Moreover, the majority of work on L2 tone acquisition tends to investigate perceptual listening performance, but in order to truly understand the learning process and attainment, tone acquisition should also be examined in other domains, such as word-learning. This paper aims to provide new insights into how linguistic and non-linguistic factors facilitate word-learning in a tone language.

Two groups of adult native speakers of English (a non-tone language, n=21) and Mandarin Chinese (a tone language, n=20) participated in two-day experiment. Musical experience (measured by years of practice and formal instruction per instrument, cf. [4]) and working memory were balanced across both groups. Participants were first tested on non-lexical tone perception through a tone identification task. They were then trained by means of a listen-and-repeat session to learn a set of 16 pseudolanguage words with a four-way segmental and four-way tonal contrast. After training, they were tested on their word recall with picture-matching. The training and the word recall task were repeated on day 2.

Overall, English and Mandarin speakers performed similarly in tone identification and word recall. A mixed regression showed that in general, only the non-linguistic factors (tone identification performance, musical experience, working memory) strongly predicted tone word acquisition. However, linguistic factors did modulate the acquisition in more subtle ways. For instance, Mandarin speakers depended less on musical background and tone perception to learn the tone words than did the English speakers, suggesting that L1 tone experience made learning tone words more like a general word-learning task. In addition, it appeared that there were transfer effects from L1 phonology, as Mandarin speakers had more difficulty than English speakers with specific tone contrasts that do not exist in Mandarin phonology. The findings of this study show a

dynamic interplay of linguistic and non-linguistic factors in L2 acquisition of tone, highlighting the importance of looking at language-specific and domain-general characteristics of second-language learners.

Bibliography

- [1] M. Antoniou and J. L. L. Chin, “What can lexical tone training studies in adults tell us about tone processing in children?,” *Front. Psychol.*, vol. 9:1, pp. 1–11, 2018. [2] A. Cooper and Y. Wang, “The influence of linguistic and musical experience on Cantonese word learning,” *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4756–4769, 2012. [3] S. Hutka, G. M. Bidelman, and S. Moreno, “Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music,” *Neuropsychologia*, vol. 71, pp. 52–63, 2015. [4] A. R. Bowles, C. B. Chang, and V. P. Karuzis, “Pitch Ability As an Aptitude for Tone Learning,” *Lang. Learn.*, vol. 66, no. 4, pp. 774–808, 2016.
-

21. Time for change: Evaluating models of semantic change without evaluation tasks

Haim Dubossarsky, Simon Hengchen, Nina Tahmasebi, and Dominik Schlechtweg
Department of Theoretical and Applied Linguistics Section, University of Cambridge

Computational models of Lexical Semantic Change (LSC) are becoming increasingly popular in language change research, and serve both for LSC detection (Bamler and Mandt 2017; Kim et al. 2014; Kulkarni et al. 2015) and for its linguistic interpretation (Dubossarsky et al. 2015; Hamilton et al. 2016; Karjus et al. 2018). However, this emerging research field lacks proper evaluation tasks, which at the very minimum should include a validated list of words that underwent semantic change in addition to semantically stable words. This lack of evaluation impedes a meaningful comparison between different models and the validation of their ability to truly capture LSC (Dubossarsky et al. 2017).

In this work we introduce two principled ways to evaluate and compare between different LSC models in the absence of a proper evaluation task. 1) We randomly assign sentences to time bins, thus effectively creating a condition in which no LSC exists (the usage of all the words is the same across all times). We maintain that any LSC that is reported under this condition cannot be attributed to REAL semantic change, and that a TRUE semantic change can only be reported if the results on the natural corpus significantly exceeds the results of the random corpus. 2) We simulate LSC by holding out a set of seed words from a natural corpus, and then replacing them with existing target words and gradually inserting them back to the corpus. In this way we mimic a new sense emergence as it unfolds over these target words.

We tested existing models of LSC using our framework and found that state-of-the-art models suffer from large amounts of noise that is associated with vector space alignment - the specific method used to facilitate the diachronic comparison of words across time. In other words, previously reported results of LSC models are likely to be related to noise factors rather than to genuine LSC.

We propose Temporal Referencing, a new model for LSC that does not require alignment. Using our novel evaluation framework, we empirically tested the performance of our Temporal Referencing model and found that it surpasses state-of-the-art models in

both its reduced noise levels and detection rates of simulated LSC. We thus propose that models' evaluation is not only possible in cases where evaluation tasks are missing, but also quite meaningful.

References:

- Bamler, R., & Mandt, S. (2017). Dynamic Word Embeddings via Skip-Gram Filtering. In ICML.
- Dubossarsky, H., Grossman, E., & Weinshall, D. (2017). Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. In EMNLP (pp. 1147–1156).
- Dubossarsky, H., Tsvetkov, Y., Dyer, C., & Grossman, E. (2015). A bottom up approach to category mapping and meaning change. In V. Pirrelli, C. Marzi, & M. Ferro (Eds.), *Word Structure and Word Usage. Proceedings of the NetWordS Final Conference* (pp. 66–70).
- Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In ACL (pp. 1489–1501).
- Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2018). Quantifying the dynamics of topical fluctuations in language, 1–29.
- Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., & Petrov, S. (2014). Temporal Analysis of Language through Neural Language Models. In ACL, (pp. 61–65).
- Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically Significant Detection of Linguistic Change. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 625–635). *Computation and Language; Information Retrieval; Learning*.

22. Towards an automated event semantic approach to L2 learner verb inflection errors

Carlos Balhana

Department of Theoretical and Applied Linguistics Section, University of Cambridge

Error analysis (EA) of second language (L2) learner outputs can provide a structured method for categorising which non-target forms are conditioned by first language (L1) interference and underspecification in the interlanguage grammar. However, manually annotating individual error tokens at a large scale is a time-intensive process potentially subject to research bias and underrepresentative sampling (Mizumoto et al. 2012). Specifically, within the domain of tense and aspect errors, annotations should ideally capture the full range of lexical and morphosemantic inflectional features for both learner-produced (surface) error forms and reconstructed (target) forms. The current research effort also argues that there is significant explanatory and predictive value in representing the temporal sequence of events produced in a non-linear textual order, especially in learners' responses to narrative tasks. We also maintain that cross-sentential event semantic representations can boost performance of grammatical error correction (GEC) systems, and that EA and GEC processing can and should complement one another representationally and procedurally (Han et al. 2010; Sidorov et al. 2013; Zhang & Houfeng, 2014; Rozovskay & Roth 2016).

To address these challenges, we present our research on the development and implementation of two natural language processing (NLP) applications: The first, the Learner Event Annotation and Correction (LEvAnCor) framework, is a multi-tiered

annotation implementation which supports fine-grained tagging of distinct verbal phenomena (e.g. main predication, aspectual periphrases, light verb constructions, etc.), surface error and target inflectional features, lexical aspect, and temporal ordering relations across the entire text. We present the results of a gold-standard error annotation effort, whereby a total of 800 scripts produced by L2 English learners from Brazilian Portuguese, Chinese, Russian, and Arabic L1 groups were annotated using data from the EFCAMDAT corpus (Geertzen, Alexopoulou, & Korhonen 2013).

We also present an implementation of the TARSQI toolkit to process L2 learner texts, which automatically annotates event features of the verb and inflectional complexes and temporal ordering of event-denoting verb phrases. We discuss how the annotations from LEvAnCor and TARSQI can support improvements on automated classification performance of L2 learner errors. Finally, we outline some future directions for deriving robust representations of L2 development trajectories given the integration of these two systems, which we believe can benefit cross-cutting research within the broader second language acquisition (SLA) research community.

References:

- Geertzen, J., Alexopoulou, T., & Korhonen, A. (2013). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In Proceedings of the 31st Second Language Research Forum (SLRF).
- Han, N. R., Tetreault, J., Lee, S. H., & Ha, J. Y. (2010). Using error-annotated ESL data to develop an ESL error correction system. In Proceedings of LREC. Mizumoto, T., Hayashibe, Y., Komachi, M., Nagata, M., & Matsumoto, Y. (2012, December). The effect of learner corpus size in grammatical error correction of ESL writings. In Proceedings of COLING 2012: Posters (pp. 863-872). Rozovskaya, A., & Roth, D. (2016). Grammatical error correction: Machine translation and classifiers. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 2205-2215).
- Sidorov, G., Gupta, A., Tozer, M., Catala, D., Catena, A., & Fuentes, S. (2013, August). Rule-based system for automatic grammar correction using syntactic n-grams for English language learning (12). In Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task (pp. 96-101).
- Zhang, L., & Houfeng, W. (2014). A unified framework for grammar error correction. In Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task (pp. 96-102).